Bioinformatics Application Note

A Handy Script to Implement Genetic Models in R

Wei Zhang*

Section of Hematology/Oncology, Department of Medicine, The University of Chicago,

Chicago, IL 60637, USA

* Correspondence: Dr. Wei Zhang, 920 E 58th Street Room 515, The University of Chicago, Chicago, IL 60637, USA. E-mail: wzhang1@uchicago.edu

Key Words: Genotype; Genetic models; R script

Background

Genome-wide associations between genotypic and phenotypic data often assume certain genetic models [1]. A handy script in R was written to implement three commonly used genetic models: additive, dominant and recessive models. This script can be easily inserted into more complicated programs to facilitate high-throughput association studies.

Methodology

Assuming that a single nucleotide polymorphism (SNP) has genotypes of AA, AB and BB, the genotypes are coded as AA=0, AB=1 and BB=2 in an additive model; AA=0, AB=1 and BB=1 in a dominant model; and AA=0, AB=0 and BB=1 in a recessive model. The missing data can be represented using 'NA'. The source code was implemented in R, an open source statistical package [2].

How-to

Copy and paste the source code in a text file (eg. models.R). Load models.R in the R editing environment. Suppose you have a vector of genotypic data, **x** [AA, AA, BB, AB, BB, AA…]. To transform the genotypic data, simply use the following commands. '>' indicates input from keyboard.

```
>ADD(x)       #additive model
0 0 2 1 2 0 …
>DOM(x)       #dominant model
0 0 1 1 1 0 …
>RES(x)       #recessive model
0 0 1 0 1 0 …
```

To do a linear regression test between a genotypic data **x** and a phenotypic vector **y**, use the following commands.

```
>lm(y~ADD(x))       #additive model
>lm(y~DOM(x))       #dominant model
>lm(y~RES(x))       #recessive model
```

lm is a R library for linear modeling. To use the script in a high-throughput study (eg. microarray), simply insert it in an iterative procedure.

Source Code in R

```
ADD <-function(x)
{
        return(as.numeric(factor(x))-1)
}

DOM<- function(x)
{
         return (ifelse(ADD(x) > 0, 1, 0 ) )
}

RES <- function(x)
{
        return (ifelse(ADD(x) > 1, 1, 0 ) )
}
```


References

1. Ramirez J, Mirkov S, Zhang W, Chen P, Das S, Liu W, Ratain MJ, Innocenti F. Hepatocyte nuclear factor-1 alpha is associated with UGT1A1, UGT1A9 and UGT2B7 mRNA expression in human liver. The Pharmacogenomics Journal. Epub April 17, 2007.

2. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org. 2006